MOLECULAR AND CELLULAR BIOLOGY, Jan. 1994, p. 116–127
0270-7306/94/$04.00+0

Vol. 14, No. 1

DNA Sequence Requirements for Transcriptional Initiator Activity in Mammalian Cells

RAMIN JAVAHERY, ANITA KHACHI, KIERSTEN LO, BEATRICE ZENZIE-GREGORY, AND STEPHEN T. SMALE*

Howard Hughes Medical Institute, Molecular Biology Institute, and Department of Microbiology and Immunology, University of California, Los Angeles, School of Medicine, Los Angeles, California 90024-1662

Received 6 May 1993/Returned for modification 27 July 1993/Accepted 30 September 1993

A transcriptional initiator (Inr) for mammalian RNA polymerase II can be defined as a DNA sequence element that overlaps a transcription start site and is sufficient for (i) determining the start site location in a promoter that lacks a TATA box and (ii) enhancing the strength of a promoter that contains a TATA box. We have prepared synthetic promoters containing random nucleotides downstream of Sp1 binding sites to determine the range of DNA sequences that convey Inr activity. Numerous sequences behaved as functional Inrs in an in vitro transcription assay, but the Inr activities varied dramatically. An examination of the functional elements revealed loose but consistent sequence requirements, with the approximate consensus sequence Py Py A_{+1} N T/A Py Py. Most importantly, almost every functional Inr that has been described fits into the consensus sequence that we have defined. Although several proteins have been reported to bind to specific Inrs, manipulation of those elements failed to correlate protein binding with Inr activity. The simplest model to explain these results is that all or most Inrs are recognized by a universal binding protein, similar to the functional recognition of all TATA sequences by the same TATA-binding protein. The previously reported proteins that bind near specific Inr elements may augment the strength of an Inr or may impart transcriptional regulation through an Inr.

In many mammalian promoters, a DNA sequence element called a TATA box positions the transcription start site by nucleating the formation of a preinitiation complex containing RNA polymerase II and several general transcription factors (43). TATA activity can be imparted by a wide variety of A+T-rich elements (10, 31, 42), a fact that several years ago led to the hypothesis that the various TATA sequences were recognized by multiple distinct TATA-binding proteins within a cell (5, 11, 30, 39). This model was supported by the finding that some upstream activators stimulate transcription through only a subset of the known TATA sequences (5, 11, 30, 38, 39). However, only one TATA-binding protein has been isolated, and this protein, called TBP, functionally interacts with numerous A+T-rich elements (10, 13, 42, 44). Apparently, the relatively low degree of sequence selectivity results from the fact that TBP recognizes the minor groove of DNA (16, 36). Because TBP is now thought to interact with every TATA box, several alternative hypotheses have been proposed to explain the functional differences among TATA elements (32). One hypothesis is that TBP interacts with the various TATA sequences in conjunction with different sets of accessory factors or with proteins that recognize overlapping control elements (1, 2, 8, 23, 24, 37).

In some promoters, a second type of core promoter element, called an initiator (Inr), has been found to mediate the same functions as TATA elements (reviewed in references 15, 32, and 40). In the narrow definition used here, Inr elements can localize a transcription start site and mediate the action of at least some upstream activators in the absence of a TATA box. An Inr carries out this activity from a position that directly overlaps the start site. Inr elements also can cooperate with a TATA box if both elements are found in the same core promoter and are separated from each other by approximately 25 bp (21, 33, 34). A variety of proteins, including TFII-I, USF, YY1, and HIP1, have been reported to impart Inr activity by functionally interacting with unique elements found at the start sites of specific genes (7, 18, 27, 28).

Two models must be considered to explain the above-described observations concerning Inr elements. First, as suggested by the identification of multiple Inr-binding proteins (7, 18, 27, 28), several distinct classes of Inr elements may exist, with each class being recognized by a different protein. In other words, the binding of any one of several distinct proteins may be sufficient for directing RNA synthesis to begin at a precise nucleotide in the absence of a TATA box. Second, all of these Inr elements may functionally interact with a universal recognition protein, similar to the recognition of heterogeneous TATA boxes by the universal TBP. With this model, specific Inr-binding proteins (1a, 7, 18, 27, 28) would augment promoter strength and act in conjunction with the universal protein.

To distinguish between these two opposing models for Inr recognition and function, we have prepared and analyzed a series of random and mutant Inr elements. Our results reveal a loose but consistent consensus sequence, strongly suggesting that all Inr elements are recognized by a universal protein or by a single class of closely related proteins. This model is supported by additional experiments that failed to demonstrate a correlation between the binding of previously reported proteins and Inr activity.

* Corresponding author. Phone: (310) 206-4777. Fax: (310) 206-3800.

MATERIALS AND METHODS

Plasmid DNAs. Plasmids containing promoters with random Inrs were constructed by preparing oligonucleotides (Applied Biosystems, Inc.) containing the sequence 5′-ATCGATGAATTCGAGCTCGG$(N)_4$CA$(N)_8$ACGGATC CTCTAGAGTCGACC-3′, where N represents an instruction to the synthesizer to insert a random array either of pyrimidines or of all four nucleotides. Another oligonucleotide, 5′-GGTCGACTCTAGAGGATCCG-3′, was annealed to the pool of long oligonucleotides, and the *Escherichia coli* DNA polymerase Klenow fragment (Boehringer) was then used to fill in the appropriate complementary nucleotides, creating a double-stranded fragment. This fragment was cleaved with *Eco*RI and *Bam*HI and inserted into the *Eco*RI and *Bam*HI sites of plasmid III (34), containing multiple Sp1 binding sites in vector pSP72 (Promega). After transformation of competent *E. coli* with the resulting ligation mixture, plasmid DNAs were isolated from individual ampicillin-resistant colonies, and the nucleotide sequences were determined with a Sequenase kit (U.S. Biochemicals, Inc.) and an SP6 promoter primer (Promega).

Plasmids containing promoters with Inr mutations were constructed by synthesizing both strands of the desired sequence flanked by *Sac*I (*Eco*RI in some cases) and *Bam*HI sites. When multiple mutations at a given location were desired, the synthesizer was instructed to insert a pool of random nucleotides at that position. The annealed oligonucleotides were then inserted into the appropriate restriction sites in plasmid III (34) and analyzed as described above.

Plasmids for transient transfection analysis were prepared by excising the promoters from pSP72-based vectors with *Hpa*I and *Bam*HI and inserting them into plasmid pSVPyTK (33), digested with *Cla*I (for blunt ending) and *Bgl*II.

In most cases, two independent large preparations of plasmids were made and were purified by column chromatography (Qiagen). Concentrations were accurately determined by absorbance measurements and confirmed by agarose gel electrophoresis.

In vitro transcription analysis. In vitro transcription reactions were carried out with 300 ng of plasmid DNA in nuclear extracts derived from HeLa cells as described previously (33, 45). RNA products were purified and analyzed by primer extension analysis with a ^{32}P-labeled SP6 promoter primer. cDNA products were visualized by electrophoresis on an 8% denaturing polyacrylamide gel and then autoradiography in the presence of an intensifying screen. Quantitation of signals was done primarily by Phosphorimager (Molecular Dynamics, Inc.) analysis, but a small subset of quantitation was performed by densitometry.

Transient transfection analysis. COS7 cells were transfected with replication-competent plasmids in vector pSVPyTK as described previously (33, 35). Cytoplasmic RNA was isolated after 48 h and analyzed by primer extension analysis (35).

DNA-binding assays. ^{32}P-labeled probes for gel mobility shift experiments were prepared either from plasmid DNA or directly from oligonucleotides. For probes prepared from plasmids containing synthetic promoters, the plasmids were digested with *Eco*RV and *Hin*dIII. The resulting 61-bp fragments were then 5′ end labeled with T4 polynucleotide kinase (Boehringer). For probes prepared from oligonucleotides, the wild-type 53-mer oligonucleotide 5′-TGATCAG GTCCTTTCTCCCAAAATGGAGCTCTTGACACAAA GAACAGTGAACT-3′ (underlined nucleotides represent core YY1 binding site) was synthesized, as were various mutant oligonucleotides. Another oligonucleotide, 5′-AGT TCACTGTTCT-3′, was designed to be complementary to the 3′ end of the 53-mer. Each 53-mer (200 ng) was annealed to a 10-fold excess of the 13-mer. A complete, double-stranded probe was prepared by the addition of the *E. coli* DNA polymerase Klenow fragment in the presence of [α-^{32}P]dATP and unlabeled dCTP, dGTP, and TTP. Unlabeled dATP was later added to complete the DNA synthesis. Probes prepared by both methods were purified by electrophoresis on a 6% polyacrylamide gel.

Gel mobility shift experiments were performed with 5,000 cpm of labeled probes and 5 μg of crude nuclear extract prepared from HeLa cells or from RLm11 T cells. Results similar to those shown in Fig. 7B and 8 were also obtained with YY1 that had been purified through a sequence-specific DNA affinity column (data not shown). Binding reactions were carried out at room temperature for 20 min as described previously (16a). Mobility shift gels were 5% polyacrylamide in 0.5× TBE buffer (1× TBE buffer is 90 mM Tris-borate plus 2 mM EDTA). Gels were prerun for 30 min and, after loading of the samples, electrophoresis was done for 90 min at 180 V. Quantitation of the dried gels was carried out by Phosphorimager scanning.

RESULTS

In vitro analysis of Inrs containing random pyrimidines. Most functional Inr elements contain the sequence CA at −1 and +1 (where +1 refers to the transcription start site), surrounded by a few or several pyrimidines (15, 32, 40). In addition, before functional studies of Inr elements were carried out, a consensus sequence of Py_2CAPy_5 was observed at the transcription start sites of several TATA-containing promoters (3, 6, 15). Thus, to examine the range of sequences that can impart Inr activity in mammalian cells, we first asked whether any sequence matching this loose consensus sequence was highly active or whether specificity in sequence requirements could be identified.

Eighteen plasmids were constructed to contain CA flanked by four upstream and eight downstream pyrimidines, with the pyrimidines generated randomly by an oligonucleotide synthesizer (see Materials and Methods and Fig. 1B). These plasmids were derived from plasmid III (34), which contains multiple binding sites for transcription factor Sp1 42 bp upstream from the CA within the random sequences. In addition, a weak TATA box containing the sequence GATATC is present in this plasmid 25 bp upstream from the CA. This sequence imparts TATA activity about 40 to 100 times more weakly than does the consensus sequence TATAAA (44). Because TATA and Inr elements cooperate with each other during transcription initiation, we included this sequence with the idea that it might allow us to detect activity from a weaker Inr sequences than we might with a highly G+C-rich sequence at −25 (see Fig. 5).

In vitro transcription experiments were performed with the above-mentioned plasmids in nuclear extracts derived from HeLa cells, and the resulting RNAs were analyzed by primer extension analysis. The positive control was a promoter containing the Inr element from the terminal transferase (TdT) gene (P2121; Fig. 1A, lanes 1 and 12), and the negative control was plasmid R226 (fig. 1A, lanes 2 and 13; see also Fig. 2B), containing at the start site the sequence CA surrounded by an array of purines.

The results of the transcription experiments with the random-pyrimidine Inrs revealed a wide range of promoter strengths (Fig. 1A). One promoter (plasmid R2559; lane 22) was as strong as the promoter containing the TdT Inr, and

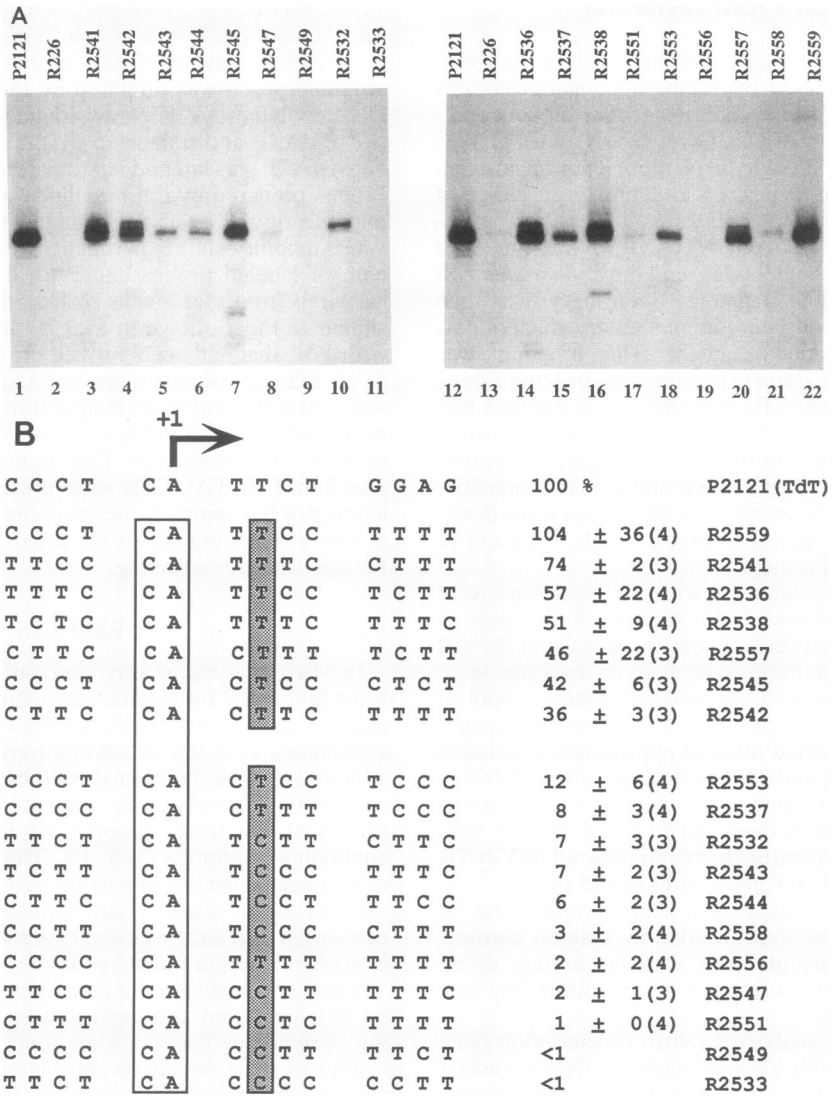


FIG. 1. Analysis of Inr elements containing random pyrimidines. Plasmids were constructed to contain promoters with Sp1 binding sites, a weak TATA box, and a potential Inr containing the sequence Py_4CAPy_8 (see Materials and Methods and Results). In vitro transcription reactions were carried out with HeLa cell nuclear extracts, and the resulting RNAs were analyzed by primer extension analysis. (A) Autoradiographs of polyacrylamide gels. The activities of the random-pyrimidine Inrs (lanes 3 to 11 and 14 to 22) are compared with the activities of the strong TdT Inr (lanes 1 and 12) and a weak Inr (R226; lanes 2 and 13; see Fig. 2). The names of the plasmids are indicated above the lanes. (B) Sequences of the Inrs and relative promoter activities, as determined by Phosphorimager analysis. The promoter activities are followed by standard deviations, which were determined from the number of independent experiments indicated in parentheses. These experiments were performed with at least two different preparations of each plasmid DNA. The open box denotes the sequence CA at −1 and +1, which was not varied in these experiments. The shaded box indicates the +3 nucleotide, which appears to be most important for determining Inr strength.

others were more than 100 times weaker (plasmids R2549 and R2533; lanes 9 and 11). Examination of the promoter strengths for the entire series revealed that numerous sequence combinations can impart at least moderate activity, with a few consistent and notable trends. Most obvious was the fact that the seven Inrs with activities of 36% or greater (relative to the TdT Inr) contain a T at +3, whereas the majority of the weaker Inrs contain a C at +3. The weakest Inr containing a T at +3, R2556, contains unusual homopolymeric stretches (C_5AT_8) that may lead to the weak activity. A second observation was that, within the sequences containing a T at +3, the stronger Inrs consistently contain a T at +2, while the weaker Inrs contain a C at +2. Also, within the sequences containing a C at +3, the stronger Inrs contain a T at +2, while the weaker Inrs contain a C at +2. These results suggest that a T at +3 may be necessary for strong Inr activity and that a T at +2 may result in a slightly stronger Inr.

A comparison of the results found with specific pairs of promoters revealed that, in some instances, sequences surrounding the −1 to +3 core can influence Inr activity. For example, plasmids R2557 and R2537 contain the same sequence between −2 and +7, yet their Inr activities differ by sixfold. In addition, plasmids R2559 and R2553 are almost

identical between −5 and +6, yet their activities differ by almost ninefold (part of this difference can be attributed to the different nucleotides at the +2 position). Finally, plasmids R2532 and R2533 are similar between −5 and +3 (except for position +2), yet their activities differ by more than sevenfold. We have not addressed the reason for the different activities found with each of these pairs of plasmids, but it may be noteworthy that in each pair, the weaker element contains unusually large numbers of G · C base pairs between −5 and +8. As mentioned above, plasmid R2556 also contains five C's preceding the start site, leading to the possibility that a high G+C content and/or G · C stretches may be detrimental for all three of the weak Inrs containing a T at +3 (R2553, R2537, and R2556). Alternative explanations for the effect of surrounding sequences are discussed below (see the Discussion).

In general, the conclusion from the experiments shown in Fig. 1 is that a wide range of sequences can impart Inr activity, with limited sequence requirements residing primarily between −1 and +3. More importantly, the consistent trend in sequence requirements observed here (CATT > CACT > CATC > CACC) suggests a model in which a single protein (or a highly related class of proteins) which binds to DNA with a low degree of sequence selectivity is responsible for the activities of all of the Inrs. If multiple, unrelated sequence-specific DNA-binding proteins were capable of imparting Inr activity, we might have expected to find a much smaller percentage of functional sequences and less similarity between the functional sequences.

In vitro analysis of Inrs containing random nucleotides. Because a wide range of sequences imparted Inr activity in Fig. 1, we extended our analysis by asking whether a similarly large proportion of sequences would be functional if the CA were surrounded by completely random nucleotides. We prepared and tested 18 plasmids in the same promoter context as that described above; the plasmids contained the sequence N_4CAN_8 at the initiation site (see Fig. 2B).

In vitro transcription analyses (Fig. 2) revealed that a smaller number of these sequences provided Inr activity and that the active Inrs, even those containing a T at +3, were less active than those containing random pyrimidines. (All start sites were combined for the quantitations presented in Fig. 2B.) The promoters with the highest activities contain either a T or an A at the +3 position and, with one exception, the weaker promoters contain either a C or a G at +3. The one exception, plasmid R247, contains a T at +3 but contains no other pyrimidines in the random sequences between −5 and +6.

With this series of plasmids, significant start site heterogeneity was observed. In most cases (e.g., R442, R450, and R452), the heterogeneity appears to have resulted from the presence of CA or TA dinucleotides at positions other than −1 and +1. However, because most of the promoters in this series are weak, it is not clear whether the presence of the additional start sites directly contributes to Inr strength.

The results of the random-nucleotide Inr analyses suggest that, in most instances, the sequence CANT/A can impart at least a low level of Inr activity. For stronger activity, a comparison of Fig. 1 and 2 reveals that pyrimidines are important in at least some positions surrounding the −1 to +3 core. Furthermore, Inr activity could be detected at low levels when a C or a G at +3 was followed by appropriate pyrimidines. However, precise sequence requirements for the surrounding pyrimidines could not be defined.

In vivo analysis of Inrs containing random sequences. The above-described experiments analyzed the activities of random-sequence Inr elements by use of an in vitro transcription assay. To test whether this set of Inrs imparts similar relative activities in vivo, we used a transient transfection assay with eight of the promoters described above. These promoters were inserted into plasmid pSVPyTK, containing a herpes simplex virus thymidine kinase reporter gene and a simian virus 40 origin of replication. This vector replicates to a high copy number in COS7 monkey cells containing an integrated copy of the simian virus 40 large tumor antigen (9). The plasmids were transfected into the COS7 cells and, after 48 h, cytoplasmic RNA was isolated and analyzed by primer extension analysis with a ^{32}P-labeled primer complementary to thymidine kinase mRNA sequences. As an internal control, equal amounts of a plasmid containing Sp1 binding sites and a TATA box were cotransfected with each test plasmid (see band labeled "C" in Fig. 3).

The results from the in vivo analysis (Fig. 3) revealed a trend in Inr strengths that was consistent with the results found with the in vitro assay, but two significant differences were apparent. First, all but one of the random-sequence Inrs resulted in greater relative promoter strengths in vivo than in vitro. For example, the weakest Inr, R226, was 10% as active as the TdT Inr in vivo (Fig. 3, lanes 1 and 8) but less than 1% as active as the TdT Inr in vitro (Fig. 2). Similarly, the R2545, R332, R442, R2532, and R2549 Inrs were 81, 46, 41, 40, and 20%, respectively, as active as the TdT Inr in vivo but 42, 24, 9, 7, and <1%, respectively, as active as the TdT Inr in vitro. Similar results were found after transfection into human 293 cells, in which the plasmids do not replicate (data not shown). Most likely, the specific promoter context is responsible for the high level of background activity and the small difference between the strongest and the weakest Inrs. In particular, the GATATC sequence at −25 and the 12 upstream Sp1 binding sites may be optimal for detecting Inr activity with the in vitro assay (see Fig. 5), but they appear to strongly diminish the importance of the Inr during the transfection experiments (see the Discussion).

The second difference found between the in vivo and in vitro assays was that one Inr, R2541, was weaker in vivo than in vitro (Fig. 3, lane 2, versus Fig. 1, lane 3), relative to the TdT Inr and to the other random-sequence Inrs. We do not understand the basis for this deviation from the trend, but one possibility is that differences in transcript stability may be more apparent with the in vivo assay. Because of the potential difficulty with transcript stability in vivo and because the in vivo assay was less sensitive in distinguishing differences between the strengths of various Inr sequences, we used the in vitro assay for the more extensive analysis described below.

Analysis of Inrs containing directed mutations. To confirm and extend the results from the random-sequence analyses, several series of promoters containing directed mutations in the TdT Inr were prepared and analyzed. We were particularly interested in establishing the sequence requirements at the −1 and +1 positions and in determining whether the elements with A or T at +3 represent two distinct classes of Inrs. The results from this analysis are shown in Fig. 4. (We previously showed that the +2 position could be varied with little effect on Inr activity [33, 34].)

These experiments confirm that the nucleotides at +1 and +3 are the most critical for determining the level of Inr activity. The +3 position appears to be the most important, because replacement of the T with G, C, or A resulted in promoter strengths that were 3, 9, and 22%, respectively, of

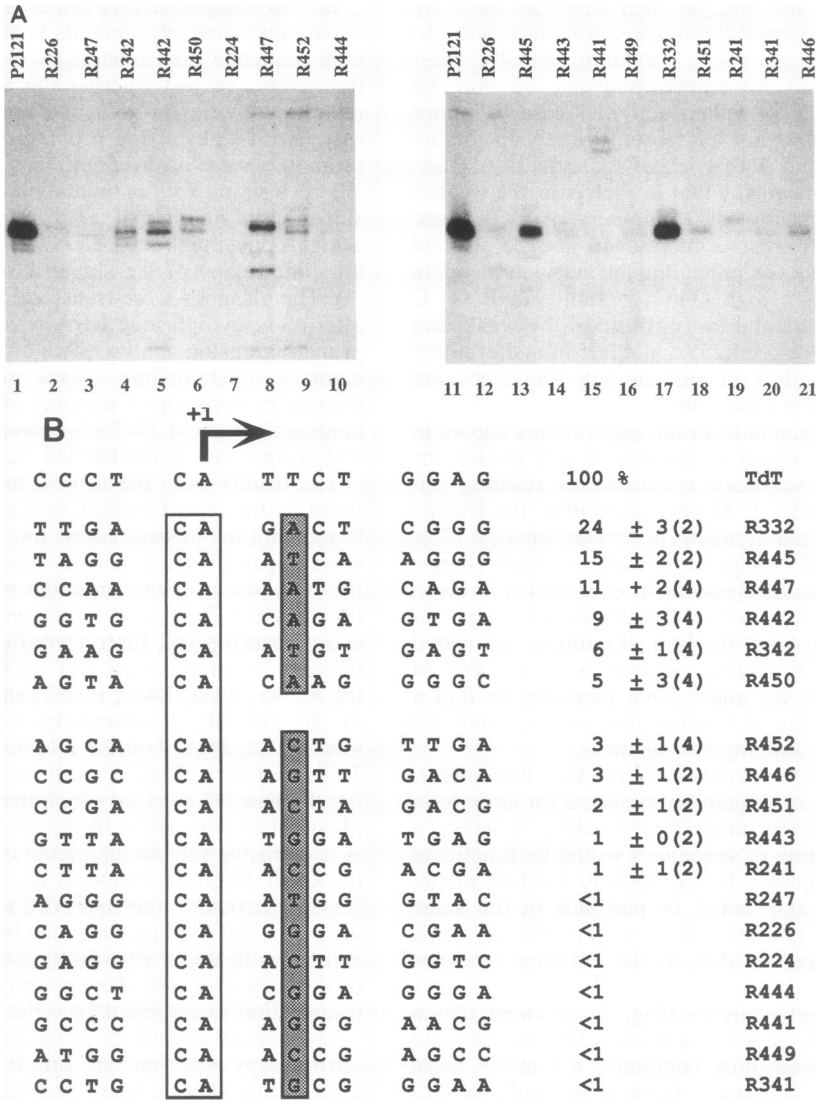


FIG. 2. Analysis of Inr elements containing random nucleotides. Plasmids were constructed and in vitro transcription reactions were performed as described in the legend to Fig. 1, except that the Inrs contained random nucleotides rather than random pyrimidines at positions −5 to −2 and +2 to +9. (A) Activities of the random-nucleotide Inrs (lanes 2 to 10 and 13 to 21) compared with the activities of the TdT Inr (lanes 1 and 11) and R226 (lanes 2 and 12). (B) Sequences of the Inrs and relative promoter activities with standard deviations, as described in the legend to Fig. 1. All start sites between approximately −5 and +5 were combined during quantitation. It should be noted that the oligonucleotide synthesizer introduced purines into the varied locations at a frequency that was greater than random. Boxes are as defined in the legend to Fig. 1.

the wild-type TdT activity (plasmids R3102, R3103, and R3121, respectively). Replacement of the A at +1 with T or G resulted in promoter strengths that were 19 and 10%, respectively, of the wild-type activity. More extensive variations of both −1 and +1 greatly reduced Inr activity (see the R1100 series). In contrast, the −5 to −2, +4, and +5 positions could be varied with relatively little effect on promoter strength. Variations at −2 or at +4 and +5 reduced promoter strength by approximately twofold, but variations at −3 to −5 had no significant effect.

The R3200 series, in which the −1 and +1 positions were varied in the presence of a constant A at +3, exhibited requirements at the −1 and +1 positions similar to those found with a T at +3 (the R1100 series). In other words, an A was still preferred at +1 and a C was preferred at −1. This result suggests that the active Inrs with an A at +3 do not represent a class of Inrs that is dramatically different from those with a T at +3.

Taken together, the results in Fig. 1, 2, and 4 show that the T at +3 and the A at +1 are the most important nucleotides for Inr activity. Pyrimidines surrounding +1 and +3 are also critical for activity, but a certain degree of redundancy exists. For example, the +4 and +5 nucleotides can be varied, with only a twofold effect, when optimal sequences are located between −5 and +3. Similarily, the −2 nucleotide can be varied, with only a twofold effect, when optimal sequences are located between −1 and +5. However, when no sequences surrounding the −1 to +3 core are optimal, the CANT sequence is insufficient for Inr activity (see R247 in Fig. 2).

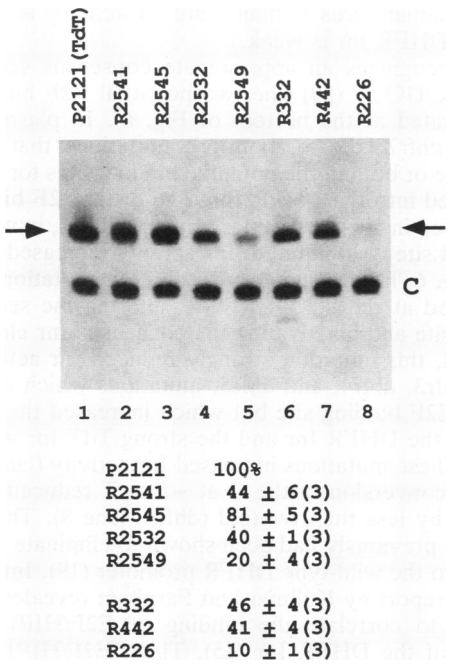


FIG. 3. In vivo analysis of Inr elements containing random sequences. Eight Inrs from the experiments in Fig. 1 and 2 were inserted into plasmid pSVPyTK. The plasmids (2 μg) were transfected into COS7 monkey cells and, after 48 h, cytoplasmic RNA was isolated and analyzed by primer extension analysis. An equal amount of a plasmid containing Sp1 binding sites and a TATA box in a similar vector was transfected as an internal control (band "C"). The names of the plasmids are listed above the lanes, and the band corresponding to the transcript from each test plasmid is indicated with an arrow. At the bottom, the relative promoter strengths are indicated, along with the standard deviations and the number of experiments used in the calculations (in parentheses).

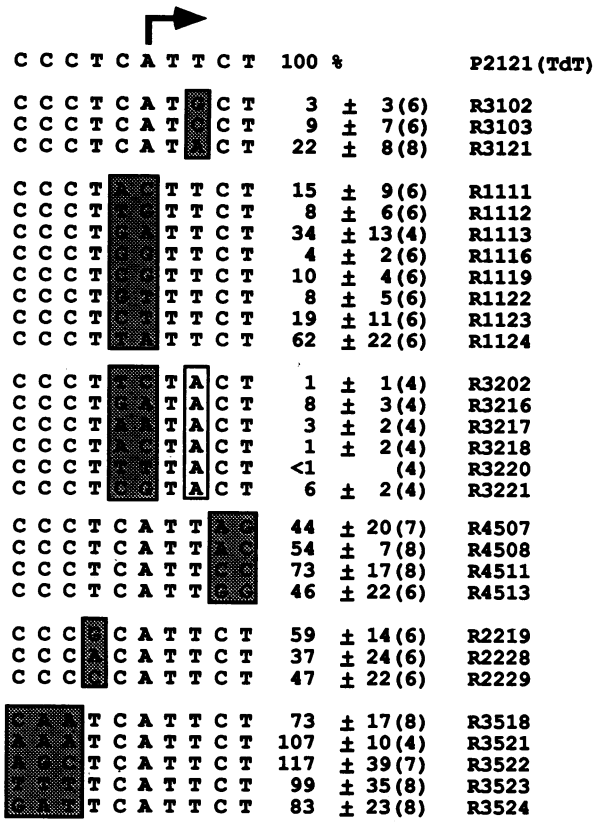


FIG. 4. Analysis of elements containing directed mutations in the TdT Inr. Plasmids were prepared to contain Sp1 binding sites, a weak TATA box, and the Inr sequences indicated and were analyzed by in vitro transcription and primer extension analyses as described in the legend to Fig. 1. The plasmid names are indicated to the right, and the shaded boxes indicate the nucleotides that were varied in each series. The open box in the R3200 series indicates that this series contains a constant A at +3. The relative promoter strengths are indicated, along with the standard deviations and the number of experiments used in the calculations (in parentheses).

Weaker Inrs depend on an A+T-rich −30 sequence for activity. In the experiments shown in Fig. 1 to 4, the promoters contained a weak TATA box (GATATC) 25 bp upstream from the random-sequence and mutant Inrs. This TATA element has an approximately 40 to 100 times lower affinity for TBP and lesser ability to function as a TATA box (44). We included this element because we suspected that it might increase the sensitivity of our assay, largely as a result of the apparent cooperativity between TATA and Inr elements. In other words, we suspected that if we instead had used a G+C-rich sequence at −25, many of the weaker Inrs that we identified would have been inactive. To test this possibility directly and to analyze the relative influences of the upstream sequence in the presence of strong versus weak Inrs, we inserted different upstream sequences into plasmids containing either the TdT Inr or a weaker Inr (containing the sequence TCTACA$_{+1}$GAGAA).

The results in Fig. 5 show that the strength of a promoter containing the TdT Inr is dependent on the sequence present between −14 and −33. This result reproduces those of experiments reported previously (44) and is most consistent with a model in which Inr-mediated transcription requires binding of TBP to the −30 region during transcription initiation. Because we can detect specific initiation with every upstream sequence, it appears that TBP can be directed to bind to the upstream region with some efficiency, regardless of the sequence present in that region (44).

In the presence of the weak Inr element, the promoter strength is also strongly dependent on the upstream sequence (Fig. 5). More importantly, with the promoters tested here, it appears that the greatest sensitivity was observed with the GATATC upstream sequence (sequence e). With this sequence, the promoter with the TdT Inr was 4.6-fold stronger than the promoter with the weak Inr, and the strength of the promoter with the weak Inr was well above background levels. In contrast, with a strong TATA box upstream (sequence a), the TdT Inr was only twofold stronger than the weak Inr. Furthermore, with upstream sequences containing a higher G+C content (i.e., a lower affinity for TBP; sequences c and d), the TdT Inr was at least five- to sevenfold stronger than the weak Inr, but the activities of the promoters containing the weak Inr were approximately at background levels. Thus, if our analysis had been performed with a stronger upstream TATA box, the differences between the strong and weak Inrs would have been smaller. If the analysis had been performed with a G+C-rich upstream sequence, the sensitivity of the assay would have been diminished because many of the weaker Inrs would have given signals near background levels. These findings also demonstrate that many of the weaker elements

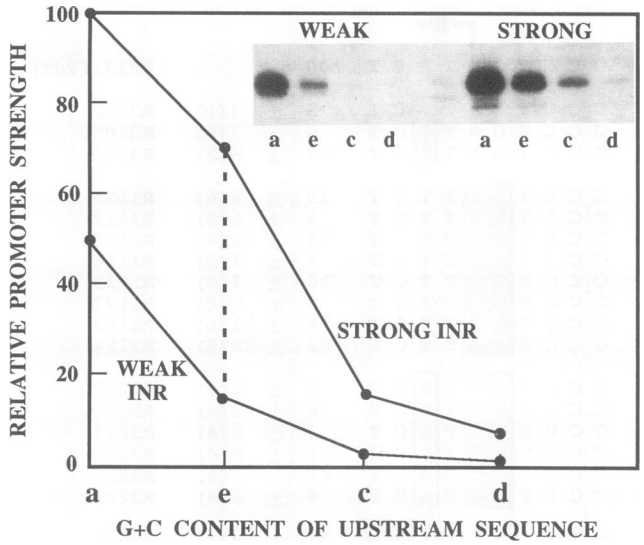


FIG. 5. Inr activity depends on the sequence present between −14 and −33. Plasmids were prepared to contain either the TdT Inr or a weaker Inr (TCTACA+1GAGAA), with one of four different sequences between −33 and −14: a (GCTATAAAAGCGATGAA TTC; 35% G+C), e (CTGATATCATCGATGAATTC; 35% G+C, but lacking the TATA consensus sequence), c (GCGTCGCCTGC GATGAATTC; 60% G+C), or d (GCGTCGCCTGGGGGTGGG GG; 85% G+C). In vitro transcription reactions were performed with 300 ng of template DNA. A representative autoradiograph is shown with the upstream sequence indicated below and the presence of a strong or weak Inr indicated above. The graph plots the relative promoter strength (as a percentage of the strongest promoter strength) versus the approximate G+C content of the upstream sequence. The broken line indicates the results found with the upstream sequence (e) that was used for the experiments shown in Fig. 1 to 4, 6, and 7.

tested in Fig. 1, 2, and 4 may not perfectly fit the definition of an Inr, because their activities probably rely on the presence of a TATA-like sequence at −30 (see the Discussion).

Analysis of Inrs which interact with E2F and YY1. The simplest explanation for the results presented above is that all active Inrs interact with a universal protein which recognizes DNA with a relatively low degree of sequence selectivity. However, previous in vitro studies suggested a functional role for at least four distinct DNA-binding proteins, YY1, E2F/HIP1, TFII-I, and USF (1a, 7, 18, 27, 28), implying instead that multiple proteins impart activity through multiple classes of Inr elements. Below we describe analyses of additional mutations which were designed to determine whether E2F/HIP1 or YY1 is responsible for the Inr activity characterized above. TFII-I and USF are discussed in the Discussion.

The dihydrofolate reductase (DHFR) gene contains an active Inr element that is recognized by a protein called HIP1 (18), now known to be similar to E2F (1a, 12). The DHFR Inr partially matches our consensus sequence and, as would be predicted from the results shown in Fig. 1 to 4, is weakly active [Fig. 6, lane 10; dhfr (WT)]. (We previously reported [21] that the DHFR Inr was strongly active in our experiments and that the simian virus 40 major late Inr was only weakly active. However, during the course of the analysis described here, we found that we had inadvertently mislabeled our plasmid DNAs during the previous study and that the simian virus 40 major late Inr actually is strong and that the DHFR Inr is weak.)

E2F recognizes an approximate consensus sequence of TTTC/GC/GCGC (20) (the two potential E2F binding sites are indicated at the bottom of Fig. 6). In plasmids dhfr1 through dhfr7 (Fig. 6, bottom), mutations that eliminate either one or both of the potential binding sites for E2F were introduced into the DHFR Inr. The distal E2F binding site was altered in all seven mutant Inrs. In dhfr1, in which only the distal site was mutated, Inr activity increased from 7 to 18% (Fig. 6, lanes 3 and 10). In dhfr2, a mutation was also introduced at the +1 nucleotide, altering the second E2F binding site and inactivating the consensus Inr element. As expected, this mutation strongly reduced Inr activity (lane 4). In dhfr3, dhfr4, and dhfr5, mutations which altered the second E2F binding site but which increased the similarity between the DHFR Inr and the strong TdT Inr were introduced. These mutations increased Inr activity (lanes 5 to 7). Finally, conversion of the G at −5 to T reduced promoter strength by less than twofold (dhfr6; lane 8). This specific mutation previously had been shown to eliminate E2F/HIP1 binding to the wild-type DHFR promoter (19). Importantly, a recent report by Kollmar and Farnham revealed a similar inability to correlate the binding of E2F/HIP1 with the activity of the DHFR Inr (15). Thus, E2F/HIP1 does not appear to be responsible for functionally interacting with the consensus Inr element.

The YY1 protein has been shown to bind to the TdT Inr and also to the Inr element in the adeno-associated virus p5 gene (28). The p5 Inr (Fig. 7, bottom, p5+1) matches closely with the weak consensus sequence defined here. Thus, YY1 is a candidate for the functional Inr-binding protein. To test this possibility, we prepared promoter constructs containing wild-type and mutant YY1 binding sites which previously had been analyzed for their abilities to interact with YY1 (22, 29). We also confirmed the ability of YY1 to bind to these elements (Fig. 7B). The p5+1 sequence, found at the start site of the p5 promoter, imparted significant Inr activity when placed downstream of Sp1 binding sites (Fig. 7A, lane 3), as reported by Seto et al. (28). Two mutations in this Inr element were tested; both had been shown by Shi et al. (29) to eliminate YY1 binding. One of these mutations (p5+1m1; Fig. 7A, lane 7) strongly reduced Inr activity, but the other (p5+1m2; lane 6) increased activity by about twofold. Next, we tested two additional sites that have been reported to bind to YY1 with a high affinity (22, 29). One of these sites, found in the upstream region of the p5 promoter (p5−60), matches our Inr consensus sequence, but the other, found in the immunoglobulin μ enhancer (E1 [22]), does not (it contains a C at the +3 position). The p5-60 element imparted strong Inr activity, but the E1 element did not (Fig. 7A, lanes 4 and 8). In addition, a mutation in the p5-60 element (p5-60m3) which previously had been found to eliminate YY1 binding (29) reduced Inr activity by less than twofold. Thus, we found no correlation between Inr activity and the binding of YY1. Instead, we found that the elements that match our loose consensus sequence impart activity, while those that diverge from our consensus sequence are largely inactive.

The ability of YY1 to bind to the wild-type and mutant Inr elements used in Fig. 7A was confirmed by a gel mobility shift assay (Fig. 7B). As expected, with a ^{32}P-labeled probe containing the p5+1 element, a strong complex was observed in nuclear extracts derived from HeLa cells (Fig. 7B, lane 2). This complex migrated at the same location as complexes detected with YY1 purified on a sequence-spe-

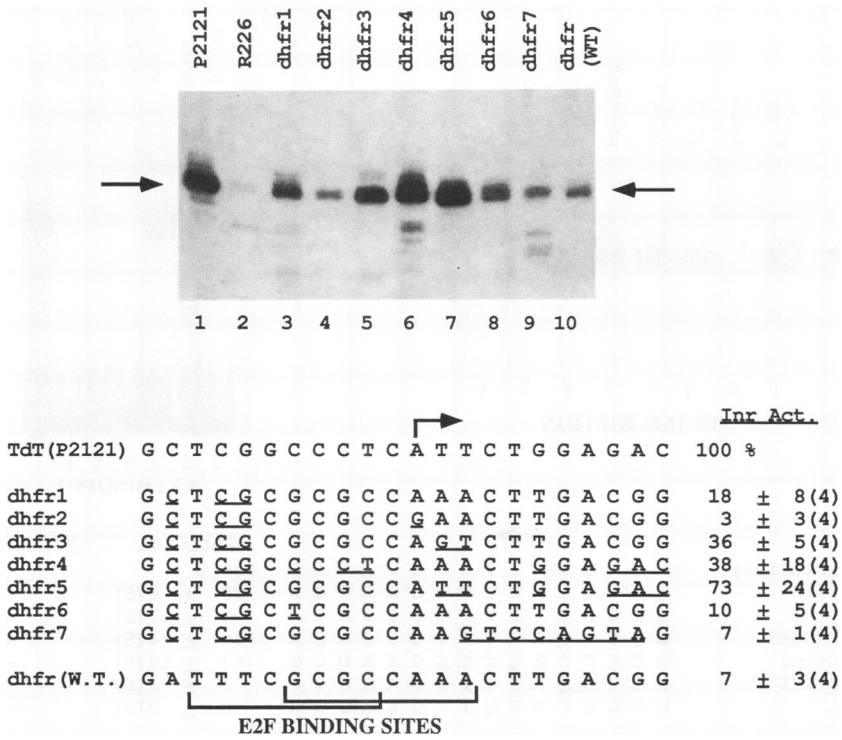


FIG. 6. Analysis of Inr elements recognized by E2F/HIP1 protein. In vitro transcription reactions were performed with plasmids containing the DHFR Inr [dhfr (WT)], which binds E2F/HIP1, or with plasmids containing mutations in the DHFR Inr. At the top, the results from a typical experiment are shown, with the name of the plasmid indicated above each lane. At the bottom, the sequences and promoter strengths (Act., activity) resulting from each plasmid are shown [with standard deviations and number of experiments used in the calculations (in parentheses)]. Specifically mutated nucleotides are underlined. The two potential E2F/HIP binding sites are indicated at the bottom (19, 20). Kollmar and Farnham (15) recently reported a similar inability to correlate the binding of HIP1/(E2F) with Inr activity.

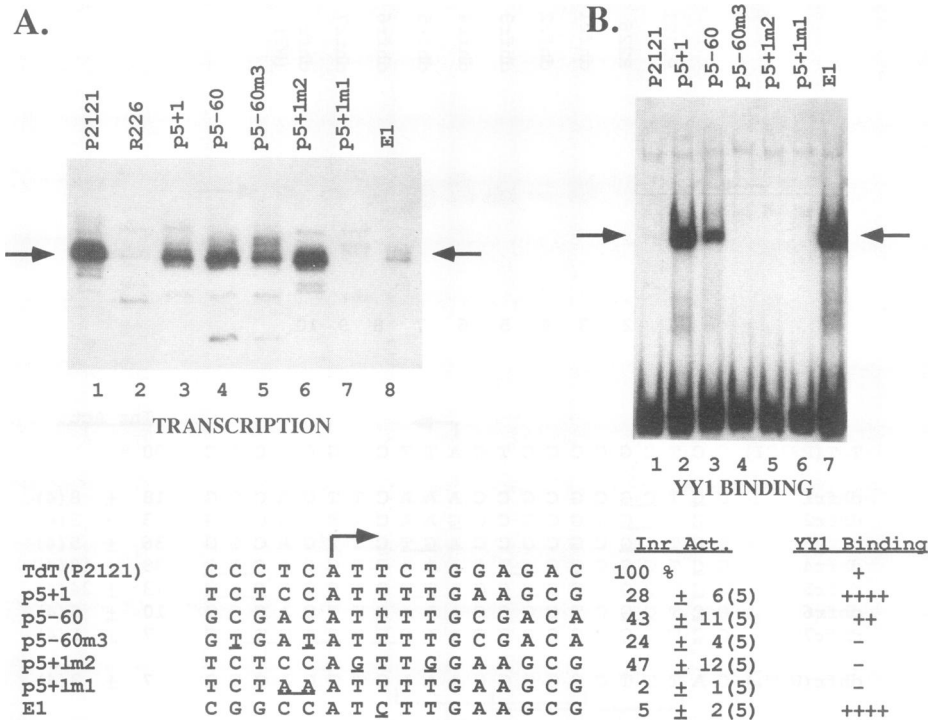
cific DNA affinity column and also with recombinant YY1 expressed in *E. coli* (data not shown). Moreover, polyclonal antibodies directed against YY1 specifically altered the migration of the protein-DNA complex (data not shown). Consistent with previous reports (22, 29), both the E1 and the p5−60 elements bound YY1 (Fig. 7B, lanes 3 and 7), although p5−60 bound with an apparently lower affinity. In some experiments, a weak band could also be detected with the TdT Inr (Fig. 7B, lane 1). However, protein-DNA complexes were not detected with the p5−60m3, p5+1m2, and p5+1m1 elements (Fig. 7B, lanes 4 to 6). These results match those reported previously (22, 29).

To confirm that YY1 does not bind to a consensus sequence matching that required for Inr activity, we defined a consensus YY1 binding site. For this analysis, we began with an oligonucleotide containing a strong YY1 binding site that is found 105 bp upstream of the transcription start site in the murine TdT promoter (17) (Fig. 8). Single-base-pair mutations were then introduced at several positions within this site, and double-stranded ^{32}P-labeled probes were prepared.

The results of the gel shift analyses (Fig. 8) defined a 4-bp core YY1 binding site (cCAT). Mutation of the first C to either a T or an A reduced the intensity of the protein-DNA complex by 3- to 4-fold (Fig. 8, lanes 2 and 4), whereas mutation to a G reduced the intensity by 25-fold (lane 3). Any mutation in the remaining three positions of the consensus sequence dramatically reduced complex formation (Fig. 8, lanes 5 to 13). In contrast, mutations in the nucleotides surrounding this core (Fig. 8, lanes 15 to 28) resulted in much smaller differences, with the strongest effect being from mutations at the T located 3 nucleotides downstream of cCAT. On the basis of these results, the sequences required for YY1 binding are clearly distinct from the consensus sequence for Inr activity.

DISCUSSION

The analysis presented in this paper suggests that all or almost all core Inr elements contain an underlying sequence similarity. However, because an unusually wide range of sequences can impart detectable Inr activity, a simple Inr consensus sequence cannot be declared. To predict the degree to which a given sequence will function as an Inr, it may be best to begin by focusing on the strong preference for an A near the start site, with a T located 2 nucleotides downstream. (As we demonstrated previously [21, 45], an Inr can enhance promoter strength even if it is shifted slightly upstream or downstream from the start site. Thus, the A can be located between approximately −5 and +5.) Next, the pyrimidine content surrounding the ANT should be assessed, with particular attention to the −1, −2, +4, and +5 positions. In general, the larger the number of pyrimidines in these locations, the greater the activity of the Inr. Moreover, if a large number of pyrimidines are present surrounding the start site, low levels of Inr activity can be imparted in the absence of either the A at +1 or the T at +3 (Fig. 4). Finally, it should be noted that previous experiments demonstrated that an Inr is unidirectional and cannot function in its reverse orientation (21).



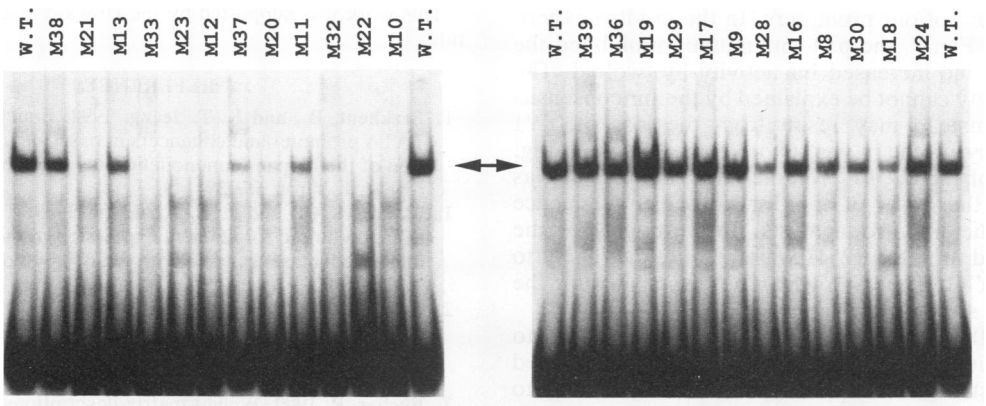
		Inr Act.	YY1 Binding
TdT(P2121)	C C C T C A T T C T G G A G A C	100 %	+
p5+1	T C T C C A T T T T G A A G C G	28 ± 6(5)	++++
p5-60	G C G A C A T T T T G C G A C A	43 ± 11(5)	++
p5-60m3	G T G A T A T T T T G C G A C A	24 ± 4(5)	-
p5+1m2	T C T C C A G T T T G G A A G C G	47 ± 12(5)	-
p5+1m1	T C T A A A T T T T G A A G C G	2 ± 1(5)	-
E1	C G G C C A T C T T G A A G C G	5 ± 2(5)	++++

FIG. 7. Analysis of Inr elements recognized by YY1. (A) In vitro transcription reactions were performed with plasmids containing a YY1 binding site at the transcription start site or with plasmids containing mutations in the binding site. The results from a typical experiment are shown, with the name of the plasmid indicated above each lane. (B) Gel mobility shift analysis of wild-type and mutant YY1 binding sites. The Inr elements from the plasmids tested in panel A were excised with *Eco*RI and *Hin*dIII and ^{32}P labeled. Gel mobility shift experiments were performed by use of 5% native polyacrylamide gels with 0.5× TBE buffer as described in Materials and Methods. At the bottom, the sequences of each Inr are shown, along with the relative promoter strengths (Act., activity) [with standard deviations and the number of experiments used in the calculations (in parentheses)] and the relative affinities for YY1. Specifically mutated nucleotides are underlined for the p5 mutations. For the E1 element, the C that leads to the weak Inr activity is underlined.

The results in Fig. 1 demonstrate that, in some instances, sequences surrounding the core Inr consensus sequence influence activity. In particular, the R2553, R2537, and R2533 Inrs are weaker than would have been predicted on the basis of the Inr consensus sequence and on a comparison with similar, randomly generated elements (see the Results). Possibly, the specific sequences found in these elements possess a reduced affinity for the putative Inr-binding protein. However, there are two alternative explanations. One alternative is based on the observation that all three of these weak Inrs contain an unusually large number of G · C base pairs between −5 and +8. Possibly, the high G+C contents reduce Inr activity by preventing template melting during transcription initiation. A second possibility is that the specific sequences contain fortuitous high-affinity binding sites for other proteins. These proteins may reduce Inr activity by preventing the binding of the functional Inr-binding protein. Further experiments are needed to explain the reason for the weak activities of the R2553, R2537, and R2533 elements.

The approximate consensus sequence Py Py A_{+1} N T/A Py Py provides a starting point for predicting whether a given sequence will impart Inr activity, but other issues must be considered when attempting to assess the importance of an Inr element within a given promoter. For example, the nature of the sequence at −25 appears to influence both the relative activity of an Inr and the degree to which the Inr influences promoter strength and start site placement (41, 44). A strong Inr, like that found in the TdT promoter, appears to function to some degree regardless of the upstream sequence (44). However, weaker Inrs may have little influence on core promoter activity and on start site placement when the promoter contains a strong consensus TATA box (Fig. 5) (17). Weak Inrs may also be inactive when the sequence at −25 is G+C rich and possesses an unusually low affinity for TBP (44) (Fig. 5). In fact, on the basis of the results shown in Fig. 5, many of the weaker elements analyzed in this study do not fit the definition of an Inr (see introductory paragraphs) in that they apparently would not be capable of dictating the location of a start site in the absence of a TATA box. Nevertheless, the weak elements most likely act through the same mechanism as the strong Inrs.

The quantitative differences between the in vitro and in vivo results further reveal that we cannot predict how well a given sequence will function as an Inr in the context of an endogenous promoter. The in vitro assay suggested that the TdT Inr was more than 100-fold stronger than the R2549 and R226 Inrs, but the in vivo assay revealed only 5- and 10-fold differences, respectively. Stable transfection assays with our synthetic promoters may help to establish the relative Inr activities in the context of a single-copy gene within a chromosome. However, even in stable transfection experiments, additional factors, such as the sequence 25 bp upstream of the start site, are likely to influence relative Inr activities. Variations in Inr strength may also result from the presence or absence of specific upstream activator elements. In light of these issues, it is best to view the results presented



W.T.	G	C	T	C	C	A	T	T	T	T	G	G	100%
M38	G	C	T	T	C	A	T	T	T	T	G	G	33
M21	G	C	T	G	C	A	T	T	T	T	G	G	4
M13	G	C	T	A	C	A	T	T	T	T	G	G	27
M33	G	C	T	C	T	A	T	T	T	T	G	G	1
M23	G	C	T	C	G	A	T	T	T	T	G	G	<1
M12	G	C	T	C	A	A	T	T	T	T	G	G	<1
M37	G	C	T	C	C	G	T	T	T	T	G	G	3
M20	G	C	T	C	C	T	T	T	T	T	G	G	2
M11	G	C	T	C	C	C	T	T	T	T	G	G	11
M32	G	C	T	C	C	A	C	T	T	T	G	G	2
M22	G	C	T	C	C	A	A	T	T	T	G	G	1
M10	G	C	T	C	C	A	G	T	T	T	G	G	1

W.T.	G	C	T	C	C	A	T	T	T	T	G	G	100%
M39	G	A	T	C	C	A	T	A	T	T	G	G	97
M36	G	A	T	C	C	A	T	T	T	T	G	G	89
M19	G	C	G	C	C	A	T	T	T	T	G	G	205
M29	G	C	T	C	C	A	T	C	T	T	G	G	71
M17	G	C	T	C	C	A	T	A	T	T	G	G	98
M9	G	C	T	C	C	A	T	G	T	T	G	G	61
M28	G	C	T	C	C	A	T	T	C	T	G	G	44
M16	G	C	T	C	C	A	T	T	A	T	G	G	98
M8	G	C	T	C	C	A	T	T	G	T	G	G	46
M30	G	C	T	C	C	A	T	T	T	C	G	G	24
M18	G	C	T	C	C	A	T	T	T	G	G	G	20
M24	G	C	T	C	C	A	T	T	T	T	T	G	57

FIG. 8. Mutagenesis of a high-affinity YY1 binding site. Double-stranded ^{32}P-labeled probes were prepared from an oligonucleotide containing a high-affinity binding site for YY1 or from oligonucleotides containing mutations in the YY1 binding site. Gel mobility shift assays were performed with 5 μg of crude nuclear extracts from HeLa cells. Similar results were found with YY1 purified on a sequence-specific DNA affinity column (data not shown). The name of the oligonucleotide is indicated above each lane, and the arrows point to the band of interest. At the bottom, the sequences in the vicinity of the YY1 binding site are indicated, and the mutated nucleotides are underlined. The relative intensities of the complexes were determined by Phosphorimager scanning of the dried gels. The percentages indicated were derived from the average of three independent experiments. W.T., wild type.

in this study as a general determination of which sequences function well as Inrs and of which sequences function less well. The precise strength and importance of a potential Inr sequence will need to be determined on a case-by-case basis and in the context of the actual promoter of interest. Most likely, similar concerns should be taken into account when assessing the strength of a potential TATA sequence.

The demonstration of an underlying similarity among most or all Inr elements is most consistent with a model in which all Inrs are recognized by a universal protein which binds to DNA with a low degree of sequence selectivity. Alternatively, the various Inr elements may be recognized by a class of Inr-binding proteins whose structures are closely related to each other. However, the results are much less consistent with previously proposed models, which suggested that Inr activity can be imparted by multiple DNA binding proteins, such as TFII-I, USF, YY1, and E2F, which are structurally very different from each other (7, 18, 27, 28, 40). In fact, none of these proteins appear to recognize the precise sequences that make up the Inr consensus sequence.

TFII-I and USF have been shown to bind with highest affinity to the sequence CACGTG, which does not match the Inr consensus sequence (27). Moreover, these proteins bind with lower affinities to the start site regions for three promoters, the TdT, adenovirus major late (AdML), and human immunodeficiency virus type 1 promoters (7, 27). The TdT and AdML Inrs match the consensus sequence described here, but the human immunodeficiency virus type 1 start site sequence (CTGGGTCT) differs significantly from the consensus sequence. The fact that TFII-I and USF appear to bind to the human immunodeficiency virus element with a similar affinity as to the TdT and AdML Inrs confirms that these proteins do not specifically recognize the Inr consensus sequence. Furthermore, we previously reported a mutational analysis which showed that the human immunodeficiency virus type 1 start site region is not functionally analogous to the TdT and AdML Inrs (21, 46). Thus, the binding of TFII-I to a start site region is not sufficient for Inr activity.

The experiments reported in this paper revealed that YY1 and E2F/HIP1 also do not bind to the Inr consensus sequence. Our results with the DHFR Inr support a recent report by Kollmar and Farnham (15), who failed to find a correlation between DHFR Inr activity and E2F/HIP1 binding. The consensus sequence that we have defined for YY1 binding reveals that this protein recognizes a core sequence, cCAT, that is distinct from the Inr consensus sequence. In an Inr which binds to YY1, the two most critical nucleotides for YY1 binding are the C at −1 and the T at +2. Conversely, Inr activity depends most strongly on the A at +1 and the T at +3. It is clear, however, that YY1 is capable of binding with high affinity to a variety of Inr elements. It remains to be determined whether YY1 is of general importance as an auxiliary protein which modulates Inr activity. Although YY1 has been reported to activate transcription through an Inr element, it is more likely to repress transcrip-

tion in the context of our promoters. In the in vitro experiments shown in Fig. 7, the p5+1m2 mutation inhibited the binding of YY1 and increased Inr activity by twofold. The increase in activity cannot be explained by the Inr consensus sequence and instead may result from the loss of YY1 binding. Similarly, in the in vivo experiments shown in Fig. 3, the activity of the R2541 promoter was lower than was anticipated on the basis of the Inr consensus sequence derived from the in vitro analysis. Interestingly, of the promoters tested in vivo, R2541 was the only one likely to bind tightly to YY1. Thus, YY1 may be responsible for the reduced in vivo strength of this Inr.

Although TFII-I, USF, YY1, and E2F are not likely to impart Inr activity, all of these proteins have been reported to enhance the strengths of specific promoters by binding to sequences which overlap transcription start sites (1a, 7, 18, 27, 28). Thus, these proteins appear to be transcriptional activators that stimulate transcription from promoters that contain either a TATA box or an Inr. Further experiments are needed to establish whether these activators function by a mechanism that is similar to or different from the mechanism used by proteins bound upstream of the core promoter. As a starting point, mutations that eliminate either the binding of a specific protein or the Inr consensus sequence (but not both) need to be introduced into the start site regions of specific natural promoters, such as the AdML, DHFR, and adeno-associated virus p5 promoters (7, 18, 28).

Because our experiments were performed with Sp1 binding sites located upstream of the Inr, we cannot rule out the possibility that different upstream activators will function through a different class of Inrs with a distinct consensus sequence. However, it should be noted that almost every Inr that has been described functions with upstream Sp1 binding sites (15, 32, 40). In addition, we found a similar consensus sequence when several of the random-nucleotide and mutant Inrs tested here were analyzed downstream of a TATA box, in the absence of an upstream activator (14). Most importantly, the consensus sequence identified here is remarkably similar to a consensus sequence derived solely through sequence comparisons (in the absence of functional studies) of over 500 unrelated genes (3). Thus, the functional sequences identified in this study are likely to be representative of the elements present in most or all Inr-containing promoters.

From the experiments presented in this study, it is apparent that a protein that recognizes the Inr consensus sequence has not been reported. However, recent studies done in our laboratory (14) and by Purnell and Gilmour (25, 26) found that the human and *Drosophila* TFIID complexes specifically interact with Inr elements. Moreover, in all of those studies, TFIID binding was strongly and specifically dependent on the Inr consensus sequence, in particular, the nucleotides at the +1 and +3 positions. A previous report by Carcamo et al. (4) revealed that the purified RNA polymerase II holoenzyme directs low levels of transcription that initiate with some specificity from CA dinucleotides. Thus, the initial recognition of the Inr consensus sequence appears to be carried out by a component of the TFIID complex. Once RNA polymerase II has been recruited to the preinitiation complex, it may preferentially begin RNA synthesis from CA dinucleotides within the Inr.

ACKNOWLEDGMENTS

We thank Doug Black, Michael Carey, and Patricia Johnson for critical reading of the manuscript and Beverly Purnell and David Gilmour for communicating results prior to their publication.

This work was supported by the Howard Hughes Medical Institute.

REFERENCES

1. **Berkhout, B., and K.-T. Jeang.** 1992. Functional roles for the TATA promoter and enhancers in basal and *tat*-induced expression of the human immunodeficiency virus type 1 long terminal repeat. J. Virol. **66**:139–149.
1a. **Blake, M. C., and J. C. Azizkhan.** 1989. Transcription factor E2F is required for efficient expression of the hamster dihydrofolate reductase gene in vitro and in vivo. Mol. Cell. Biol. **9**:4994–5002.
2. **Brou, C., S. Chaudhary, I. Davidson, Y. Lutz, J. Wu, J.-M. Egly, L. Tora, and P. Chambon.** 1993. Distinct TFIID complexes mediate the effect of different transcriptional activators. EMBO J. **12**:489–499.
3. **Bucher, P.** 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J. Mol. Biol. **212**:563–578.
4. **Carcamo, J., L. Buckbinder, and D. Reinberg.** 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc. Natl. Acad. Sci. USA **88**:8052–8056.
5. **Chen, W., and K. Struhl.** 1988. Saturation mutagenesis of a yeast *his3* "TATA element": genetic evidence for a specific TATA-binding protein. Proc. Natl. Acad. Sci. USA **85**:2691–2695.
6. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1405–1414.
7. **Du, H., A. L. Roy, and R. G. Roeder.** 1993. Human transcription factor USF stimulates transcription through initiator elements of the HIV-1 and the Ad-ML promoters. EMBO J. **12**:501–511.
8. **Dynlacht, B. D., T. Hoey, and R. Tjian.** 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell **66**:563–576.
9. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
10. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA **86**:5718–5722.
11. **Harbury, P. A., B., and K. Struhl.** 1989. Functional distinctions between yeast TATA elements. Mol. Cell. Biol. **9**:5298–5304.
12. **Kaelin, W. G., W. Krek, W. R. Sellers, J. A. DeCaprio, F. Ajchenbaum, C. S. Fuchs, T. Chittenden, Y. Li, P. J. Farnham, M. A. Blanar, D. M. Livingston, and E. K. Flemington.** 1992. Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. Cell **70**:351–364.
13. **Kao, C. C., P. M. Lieberman, M. C. Schmidt, Q. Zhou, R. Pei, and A. J. Berk.** 1990. Cloning of a transcriptionally active human TATA binding factor. Science **248**:1646–1650.
14. **Kaufmann, J., and S. T. Smale.** Direct recognition of initiator elements by a component of the transcription factor IID complex. Submitted for publication.
15. **Kollmar, R., and P. J. Farnham.** 1993. Site-specific initiation of transcription by RNA polymerase II. Proc. Soc. Exp. Biol. Med. **203**:127–139.
16. **Lee, D. K., M. Horikoshi, and R. G. Roeder.** 1991. Interaction of TFIID in the minor groove of the TATA element. Cell **67**:1241–1250.
16a. **Lo, K., N. R. Landau, and S. T. Smale.** 1991. LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. Mol. Cell. Biol. **11**: 5229–5243.
17. **Lo, K., and S. T. Smale.** Unpublished results.
18. **Means, A. L., and P. J. Farnham.** 1990. Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell. Biol. **10**:653–661.
19. **Means, A. L., J. E. Slansky, S. L. McMahon, M. W. Knuth, and P. J. Farnham.** 1992. The HIP1 binding site is required for growth regulation of the dihydrofolate reductase gene promoter. Mol. Cell. Biol. **12**:1054–1063.

20. **Nevins, J. R.** 1992. A link between the Rb tumor suppressor protein and viral oncoproteins. Science **258:**424–429.
21. **O'Shea-Greenfield, A., and S. T. Smale.** 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J. Biol. Chem. **267:**1391–1402.
22. **Park, K., and M. L. Atchison.** 1991. Isolation of a candidate repressor/activator, NF-E1 (YY-1, δ), that binds to the immunoglobulin κ 3′ enhancer and the immunoglobulin heavy-chain μE1 site. Proc. Natl. Acad. Sci. USA **88:**9804–9808.
23. **Parvin, J. D., and P. A. Sharp.** 1991. Identification of novel factors which bind specifically to the core promoter of the immunoglobulin heavy chain gene. J. Biol. Chem. **266:**22878–22886.
24. **Parvin, J. D., H. T. M. Timmers, and P. A. Sharp.** 1992. Promoter specificity of basal transcription factors. Cell **68:**1135–1144.
25. **Purnell, B. A., and D. S. Gilmour.** 1993. Contribution of sequences downstream of the TATA element to a protein-DNA complex containing the TATA-binding protein. Mol. Cell. Biol. **13:**2593–2603.
26. **Purnell, B. A., and D. S. Gilmour.** Personal communication.
27. **Roy, A. L., M. Meisterernst, P. Pognonec, and R. G. Roeder.** 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature (London) **354:**245–248.
28. **Seto, E., Y. Shi, and T. Shenk.** 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature (London) **354:**241–245.
29. **Shi, Y., E. Seto-L.-S. Chang, and T. Shenk.** 1991. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell **67:**377–388.
30. **Simon, M. C., T. M. Fisch, B. J. Benecke, J. R. Nevins, and N. Heintz.** 1988. Definition of multiple, functionally distinct TATA elements, one of which is a target in the *hsp70* promoter for E1A regulation. Cell **52:**723–729.
31. **Singer, V. L., C. R. Wobbe, and K. Struhl.** 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev. **4:**636–645.
32. **Smale, S. T.** 1994. Core promoter architecture for eukaryotic protein-coding genes, p. 63–81. *In* R. C. Conaway and J. W. Conaway (ed.), Transcription: mechanisms and regulation. Raven Press, New York.
33. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57:**103–113.
34. **Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore.** 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87:**4509–4513.
35. **Smale, S. T., and R. Tjian.** 1985. Transcription of herpes simplex virus thymidine kinase sequences under the control of wild-type and mutant human RNA polymerase I promoters. Mol. Cell. Biol. **5:**352–362.
36. **Starr, D. B., and D. K. Hawley.** 1991. TFIID binds in the minor groove of the TATA box. Cell **67:**1231–1240.
37. **Tamura, T., K. Sumita, S. Hirose, and K. Mikoshiba.** 1990. Core promoter of the mouse myelin basic protein gene governs brain-specific transcription in vitro. EMBO J. **9:**3101–3108.
38. **Taylor, I. C. A., and R. E. Kingston.** 1990. Factor substitution in a human HSP70 gene promoter: TATA-dependent and TATA-independent interactions. Mol. Cell. Biol. **10:**165–175.
39. **Wefald, F. C., B. H. Devlin, and R. S. Williams.** 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. Nature (London) **344:**260–262.
40. **Weis, L., and D. Reinberg.** 1992. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J. **6:**3300–3309.
41. **Wiley, S. R., R. J. Kraus, and J. E. Mertz.** 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the −30 region of TATA-less promoters. Proc. Natl. Acad. Sci. USA **89:**5814–5818.
42. **Wobbe, C. R., and K. Struhl.** 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol. Cell. Biol. **10:**3859–3867.
43. **Zawel, L., and D. Reinberg.** 1992. Advances in RNA polymerase II transcription. Curr. Opin. Cell Biol. **4:**488–495.
44. **Zenzie-Gregory, B., A. Khachi, I. P. Garraway, and S. T. Smale.** 1993. Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol. Cell. Biol. **13:**3841–3849.
45. **Zenzie-Gregory, B., A. O'Shea-Greenfield, and S. T. Smale.** 1992. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. J. Biol. Chem. **267:**2823–2830.
46. **Zenzie-Gregory, B., P. Sheridan, K. A. Jones, and S. T. Smale.** 1993. The HIV-1 core promoter lacks a simple initiator element but contains a bipartite activator at the transcription start site. J. Biol. Chem. **268:**15823–15832.